

基于数据库的专家系统建设

李洪磊 (山东经济学院计算机管理系 250014)

摘要:专家系统是人工智能领域中最活跃分支,也是决策支持的重要工具,但其构建通常依赖于为数较少的专用工具,如 Prolog、Lisp 等,本文作者在吸收前一时国内外众多研究成果的基础上,结合自己的工作经验和设想,提出了一个使用数据库构建专家系统的方案,以飨读者。

关键词:数据库 关系 规则 知识库 推理机

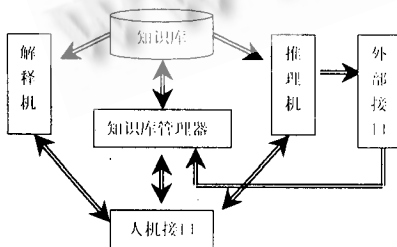
一、引言

众所周知,诞生于 70 年代的专家系统是人工智能领域最为活跃的分支,尤其在 80 年代在世界范围内更是得到了迅速发展和广泛应用。但在其发展过程中也存在一些问题阻碍其发展,其中问题之一就是开发语言专一且不易与其他程序设计语言嵌套,使以专家系统为子系统的应用系统(如 DSS)开发受到严重影响。尤其近几年来,数据仓库及相关技术的兴起使从历史数据中发现知识变成现实,但如何将管理这些知识,利用这些知识的专家系统溶于整个决策支持系统中成为摆在开发人员面前的又一难题。

由于大多数 DSS 的数据处理均依靠数据库技术实现,且目前的数据仓库也主要依靠数据库技术实现,所以采用数据库技术构建专家系统成为解决上述问题比较有效的技术手段。本文作者在吸收前一时国内外众多研究成果的基础上,结合自己的工作经验和设想,提出了以下“基于数据库的专家系统建设”方案。

二、系统设计概要

本系统整体结构如下图所示:



从上图可以看出,整个专家系统,由知识库、推理机、解释机、知识库管理器和接口(人机接口和外部接口)组成。

知识库 是知识的存储部件,是整个系统的核心,所有的操作都是围绕它进行的,因此需要单独建立一个管理器对其进行管理和维护。

知识库管理器 负责完成对知识库的管理和维护,具体工作包括:对知识库的日常维护操作(增加、删除、修改和查询)和对知识进行合法性检验。

推理机 是用于记忆所采用的规则和控制策略的程序,它使整个专家系统能够以逻辑方式协调地工作,针对用户或外部系统提出的问题根据知识库中存储的知识进行推理,给出回答,所以它也是整个系统重要部件之一。

解释机 用于向用户解释专家系统的行为,使用户能够输入数据、提出问题、了解推理过程及结果以及完成对知识库的管理等。

外部接口 实现本系统与外部系统的信息交换,一方面可从外部系统(如数据挖掘系统)获取新知识,另一方面,也可以通过该接口对外界提供服务。

三、系统详细设计

1. 知识库设计

知识库存储着大量的领域知识,这些知识由表示蕴含关系的规则表达,这些规则形式如下:

IF X Then Y 或 $X \Rightarrow Y$ 或 (X, Y)

含义为“如果 X 成立则有 Y 结论”。规则中 X 称为前项(Antecedent),是一系列条件的组合,即 $X = X_1 \wedge X_2 \wedge \dots \wedge X_n$,用以表示前提;Y 称为后项(Consequent),表示结论。有时为了表达不完全的知识,还需要引入置信度 Con,新规则形式如下:

IF X Then (Y, Con) 或 $X \Rightarrow (Y, \text{Con})$ 或 $(X, (Y, \text{Con}))$ 含义为“如果 X 成立则有 Y 结论的置信度为 Con”。

上述规则亦可表示为笛卡尔乘积的形式,即 $\langle X, Y, \text{Con} \rangle$,于是知识库可设计成如下关系:

规则序号	规则前项	结论	置信度
1	顾客为男性 购买尿布	购买啤酒	0.8
2	购买烤鸭	购买面饼、面酱	
...			

由于专家系统在使用过程中不断有新知识加入、旧知识淘汰或修改,因此为了便于对知识库中各项规则的管理,还需要将规则进行分类,并且上述关系还需进一步规范,因此最终知识库的设计结果由以下四个关系组成(以商业领域应用为例):

Rule_Category (规则种类表):存放规则种类。

种类编号	种类名称
Z001	进货
Z002	销售
Z003	存货

Rule_Antecedent (规则前项表):存放前项元(不可分割的前提条件),这些前项元可进行任意组合构成前提。

规则前项编号	规则前项
A0001	顾客为男性
Z0002	顾客为女性
A0003	购买烤鸭
A0004	购买尿布
A0005	购买电动剃须刀
...	

Rule_Consequent (规则后项表):存放结论

规则后项编号	规则后项
S0001	啤酒
S0002	面饼、面酱
S0003	电池
S0004	化妆品
...	

Rule_Precondition (规则前提表):存放规则的前提

规则序号	规则前提
1	A0001
1	A0004
2	A0003
3	A0002
3	A0004
4	A0005
...	

Rule_Conclusion (规则结论表):存放规则的结论,两者通过“规则序号”进行连接。

规则序号	前项容量	结论	置信度	规则种类
1	2	S0001	0.8	Z001
2	1	S0002		Z001
3	2	S0004	0.8	Z001
4	1	S0003	0.9	Z001
...				

注:①置信度为空时表示完全知识;②前项容量为前提中的条件数目,其作用在推理机设计中说明;③Rule_Precondition表和Rule_Conclusion表的内容共同表达了完整的规则(知识)。如知识库中规则1表示了“购买尿布的男性有80%的可能性购买啤酒”这一条规则(知识)。

2. 知识库管理器设计

知识库管理器应具备如下功能:

- (1)日常维护,包括规则的增加、修改、删除和查询
- (2)规则合法性检验,包括如下内容:

①正确性检验:根据给定的前提,依据规则能得出既定的结论,否则将该规则显示给领域专家进行判别处理。

②多义性检验:根据蕴含规则的要求,给出完备的前提应得出唯一的结论,如果知识库存在前提相同而结论不同的规则就出现了多义性错误,应将这些规则显示给领域专家进行判别处理。

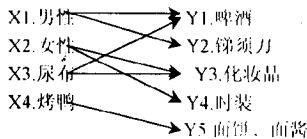
③命题包含检验:检验知识库中是否存在 $X \Rightarrow Y, X \wedge X' \Rightarrow Y$ 的情况,如存在则将这些规则显示给领域专家进行判别处理。

④命题冗余检验:检验知识库中是否存在成对的结论相同而前提中除有些条件互逆而其余条件相同(或等价)的规则,如存在则需将这些规则进行合并,消除互逆条件。

为了保证知识库中不存在非法的规则,上述检验应在新规则加入时生效,将错误消除在萌芽阶段。以上这些检验程序核心部分都由SQL语句实现,由于篇幅所限在此不作介绍。

3. 推理机设计

推理机的设计与知识库结构有直接的关系,由于知识库采用关系模式构建,所以推理机可用关系查询语句(SQL)来制作。为了详细说明推理机的设计原理,本文设计了如下关联销售知识库模型:



在该知识库模型中共有五条规则,分别描述了五条知识(以笛卡尔乘积方式表达),分别是:

$K1 = (X1, X3, Y1)$; $K2 = (X1, Y2)$; $K3 = (X2, X3, Y3)$;

$K4 = (X2, Y4)$; $K5 = (X4, Y5)$ 。

基于 $X1$ 的结论集合 $Y_{X1} = \text{Select Rule_Consequent. 规则后项 from Rule_conclusion, Rule_Precondition, Rule_CConsequent Where Rule_Precondition. 规则前提} = X1 \text{ AND Rule_Precondition. 规则序号} = \text{Rule_conclusion. 规则序号 AND Rule_conclusion. 结论} = \text{Rule_Consequent. 规则后项编号}; = \{Y1, Y2\}$ 。同理,

基于 $X2$ 的结论集合 $Y_{X2} = \{Y3, Y4\}$;

基于 $X3$ 的结论集合 $Y_{X3} = \{Y1, Y3\}$;

基于 $X4$ 的结论集合 $Y_{X4} = \{Y5\}$;

于是 $K1 = (X1, X3, Y_{X1} \wedge Y_{X3})$;

$K3 = (X2, X3, Y_{X2} \wedge Y_{X3})$;

$K5 = (X4, Y_{X4})$;

但 $K2 \neq (X1, Y_{X1})$ 、 $K4 \neq (X2, Y_{X2})$, 原因见推理机设计部分。

因此推理机设计算法如下:

$X = [x_1, x_2, \dots, x_n]$ /* 提出推导前提

$S = \text{"Declare Con Cursor for Select Rule_Consequent. 规则后项 from Rule_conclusion, Rule_Precondition, Rule_Consequent Where Rule_Precondition, 规则前提} = X_i \text{ AND Rule_Precondition. 规则序号} = \text{Rule_conclusion. 规则序号 AND Rule_conclusion. 结论} = \text{Rule_Consequent, 规则后项编号"} /* 求 $Y_{X1}$$

For $i = 2$ to N Do /* 构造 $Y_{x2} \wedge Y_{x3} \wedge \dots \wedge Y_{x_n}$ 结论推导语句

$S = S + \text{"Intersection Select Rule_Consequent. 规则后项 from Rule_conclusion, Rule_Precondition, Rule_Consequent Where Rule_Precondition. 规则前提} = X_i \text{ AND Rule_Precondition. 规则序号} = \text{Rule_conclusion. 规则序号 AND Rule_conclusion. 结论} = \text{Rule_Consequent. 规则后项编号"}$

End Do

Connect to KB /* 连接知识库

Exce SQL S /* 执行结论推导语句

Open S

Fetch S Into C

Output C /* 输出结论

Close S

算法说明:①算法设计采用了无回溯的集合操作;②上述算法能借助 SQL 语言快速完成推导过程,但当知识库中存在形如“ $X1 \Rightarrow Y1, X1 \wedge X2 \Rightarrow Y2$ ”这种“前提包含,结论不同”的多个规则时推导就会出错,如当前提为“女性”时,依据规则正确结论应为“时装”,但上述推导结论却是{时装、化妆品},显然这是错误的。为此在知识库中加入“前项容量”,并在上述算法 Select 语句的 Where 子句中加入“AND Rule_Conclusion. 前项容量 = N”解决这一问题。

4. 解释机设计

解释机应具有如下功能:

(1)解释推理的正确性

(2)解释输出其他候选解的原因

可采用显示推理依据的规则的方式实现,因解释机在整个系统中处于次要位置和篇幅所限,具体算法在此不作介绍。

5. 人一机接口设计

人一机接口又称人机界面,在设计过程中应注意做到界面友好,可采用目前流行的基于 GUI 的可视化编程工具制作。

6. 外部接口设计

外部接口的设计工作重点在于如何实现与外部系统的数据格式转化及提高通信速度,设计应遵循数据耦合原则,以避免对系统运用的影响。

四、结束语

利用数据库技术构建专家系统是一种应用性的探索,仍需要在实践中不断发展、完善,因此本文所提出的设计思想仍会有不足之处,希望能得到同仁、师长的批评、指正。

主要参考文献

- [1] 《数据库系统概念》萨师焯,王珊,高等教育出版社
- [2] 《人工智能及其应用》蔡自兴,徐光佑,清华大学出版社
- [3] 《专家系统原理与设计》张文星等,武汉测绘科技大学出版社

(来稿时间:1999年9月)